DISCUSSION PAPER SERIES

DP16325

Reduced R&D Investments: A Flip Side of Immigration?

Torje Hegna and Karen Helene Ulltveit-Moe

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

CEPR

ISSN 0265-8003

Reduced R&D Investments: A Flip Side of Immigration?

Torje Hegna and Karen Helene Ulltveit-Moe

Discussion Paper DP16325
Published 05 July 2021
Submitted 02 July 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Reduced R&D Investments: A Flip Side of Immigration?

Abstract

This paper investigates the impact of a large immigration shock on firms' R&D investments. We examine the impact of the sudden influx of immigration to Norway after the Eastern enlargement of the European Union, and the extent to which the increased immigration penetration contributed to the observed overall decline in R&D in the business sector. To identify the effect of immigration we exploit the fact that firms differ in their occupational mix and that the exposure to increased immigration varies across occupations. In our empirical approach we exploit that language requirements are significant barriers for foreign workers, and that these requirements vary across occupations. The results suggest that the immigration shock had a negative impact on firms' R&D investments with potential long term effects on productivity growth.

JEL Classification: O30, F22, F15

Keywords: R&D Investment, Immigration, productivity

Torje Hegna - t.m.hegna@econ.uio.no
University of Oslo

Karen Helene Ulltveit-Moe - k.h.ulltveit-moe@econ.uio.no
University of Oslo and CEPR

Acknowledgements
We would like to thank Kjetil Storesletten, Swati Dhingra, Andreas Moxnes and seminar and workshop participants for helpful comments and valuable discussion. This paper is part of research programs sponsored by grants 227072 and 270772 of the Research Council of Norway. Administrative registers and survey data made available by Statistics Norway have been essential.

Reduced R&D Investments: A Flip Side of Immigration?*

Torje Hegna[†] Karen Helene Ulltveit-Moe[‡]

July 2021

Abstract

This paper investigates the impact of a large immigration shock on firms' R&D investments. We examine the impact of the sudden influx of immigration to Norway after the Eastern enlargement of the European Union, and the extent to which the increased immigration penetration contributed to the observed overall decline in R&D in the business sector. To identify the effect of immigration we exploit the fact that firms differ in their occupational mix and that the exposure to increased immigration varies across occupations. In our empirical approach we exploit that language requirements are significant barriers for foreign workers, and that these requirements vary across occupations. The results suggest that the immigration shock had a negative impact on firms' R&D investments with potential long term effects on productivity growth.

*We would like to thank Kjetil Storesletten, Swati Dhingra, Andreas Moxnes and seminar and workshop participants for helpful comments and valuable discussion. This paper is part of research programs sponsored by grants 227072 and 270772 of the Research Council of Norway. Administrative registers and survey data made available by Statistics Norway have been essential.

[†]University of Oslo, torje.hegna@econ.uio.no

[‡]University of Oslo & CEPR, k.h.ulltveit-moe@econ.uio.no

1 Introduction

R&D is a key driver of productivity growth. Understanding the role of economic integration for firms' incentives to invest in R&D remains a key questions in economics. There is an extensive literature that investigates the role of trade liberalization for innovation and technical change (see Shu and Steinwender, 2019). But trade is just one aspect of economic integration. Another aspect of integration which may affect firms' incentives to invest in R&D, but which has received relatively little attention, is immigration. The scant literature on the relationship between innovation and immigration has focused on the impact of high-skilled immigrants on innovation, see e.g. Stuen et al. (2012).[1] Theory does not provide an unambiguous answer to the question of how an immigration shock may be expected to affect R&D, but seminal contributions by e.g. Acemoglu (1998) points to the relationship between labor supply and technical change. Hence, whether, and to what extent, immigration has an impact on firms' investments in R&D remains an empirical question. Based on a natural experiment we aim to add to this literature by examining how firms adjust to a significant immigration shock of relatively unskilled labor through their investments in R&D, with potential implications productivity growth.

Our empirical analysis is motivated by the fact that in the aftermath of a large and sudden immigration shock, R&D investment in the business sector in Norway declined. Over the same period, R&D investment for the group of OECD countries increased.[2] Hence, the development in Norway differs from many neighboring countries, and our hypothesis is that the immigration shock contributed to the decline in R&D investments.

Following the Eastern enlargement of the EU in 2004, which extended the common European labor market to include roughly 100 million individuals from the EU accession countries, Norway experienced a large wave of immigrants. With real wages among the highest, and unemployment among the lowest, in Europe, Norway became a popular destination for labor migrants.[3] Over the next four years the number of labor migrants quadrupled and ten years later the share of migrants in employment had risen from 7 to 17 percent. In addition to the sheer magnitude of the immigration shock, the Norwegian case is particularly useful to study since the policy change was unambiguously exogenous. As EEA-member Norway is

[1]Notable exceptions include Lewis (2011) and Gray et al. (2020).

[2]See OECD Main Science and Technology Indicators 2021, downloaded at https://stats.oecd.org/. R&D investments are measured as share of GDP.)

[3]Norway is a member of the European Economic Area (EEA) and therefore part of the EU single market. Before 2004, accession country citizens had very limited access to the Norwegian labor market. A limited number of work permits were provided via domestic employers in need of specialist competence, or on a temporary 3-month seasonal basis, typically for agricultural work. After 2004 Norway had few transitional restrictions on immigration from the accession countries compared to most EU countries (Dølvik and Eldring, 2008).

part of the single market, but not a member of the EU. Norway is thus bound to adopt EU legislation without representation in the European Parliament and Commission. The policy change was instant, comprehensive and externally imposed, providing a unique setting to study the impact of immigration.

Our empirical approach relies on two key observations: First, the sudden influx of labor to Norway varied substantially across occupations, with some occupations receiving many immigrants, while others where hardly affected. Second, even within industries firms differ in the occupational mix of their labor force. One may therefore expect that they are differently affected by the immigration shock. Using high-quality and detailed administrative Norwegian employer-employee data combined with survey data on firms' R&D investments we use the variation in exposure to the shock across firms to identify the impact of the immigration shock on R&D. An important advantage of our firm-level approach is that we are able to sweep out all industry trends in innovation by fixed effects.

Our methodology overcomes a set of empirical challenges. A common concern when analyzing the impact of immigration relates to the endogeneity of new immigrants choice of occupation and location. Immigrants may typically sort in to occupations based on current and future demand shocks. We resolve this endogeneity problem by using a shift-share instrument. Recognizing that the number of pre-enlargement immigrant workers from the EU accession countries was very low, this rules out using the standard "ethnic enclave" design (see Altonji and Card, 1991). Rather than relying on past settlements of migrants to instrument for the immigration shock, we build on Bratsberg et al. (2019) and exploit the fact that language requirements constitute significant barriers for foreign workers, and that these requirements typically vary across occupations (see e.g. Peri and Sparber, 2009). Combining occupation specific information on language requirement with firms' occupational mix in the pre-shock period, we compute a firm specific language intensity which we use to instrument for the immigration shock exposure. This exogenous occupation characteristics turns out to be a powerful predictor of immigration flows across occupations

Our choice of instrument also appears attractive in light of an important characteristic of the survey data on R&D. The stratification variables used for the R&D surveys undertaken in the EU and in Norway (Community Innovation Surveys) are industries and enterprise size according to number of employees. While regional aspects are taken into account, this is only done at a very aggregated level (NUTS2).[4] This implies that the commonly used "ethnic enclave" approach is less suitable for the analysis of the impact of immigration on R&D.

[4]https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02004R0753-20080101&from=EN#tocId10

Our results show that the immigration shock had a significant negative impact on firms' investment in R&D. The most exposed firms reduced their R&D investments in the aftermath of the immigration shock. The exclusion restriction of the instrument is violated if language intensity is associated with growth in R&D investments for other reasons than via migrant labor supply. We perform two sets of robustness checks. First, we control for pre-shock firm characteristics, specifically R&D intensity and firm size. We also add controls for R&D intensity at the industry level and for unobserved industry trends by including fixed effects. Second, if language intensity is systematically related to growth in R&D investments, even in the absence of immigration, then we should expect a significant association between R&D and language intensity for other time periods and particular for the period *before* 2004. Reassuringly, we find no such relationship. Finally we also address the medium to long term impact on productivity. We find that the immigration shock had a significant negative impact on productivity in the medium to long run. Although our analysis does not allow us to conclude on a causal relationship between the immigration induced decline in R&D investments and the reduced productivity growth, the results do indicate that there is a link.

Our paper adds to the literature on the effects of immigration on the economy through firm level adjustments. Closest in the spirit to our paper are the contributions by Lewis (2011) and Dustmann and Glitz (2015), and Gray et al. (2020). Lewis shows how supply of low-skilled immigration reduces firms growth in capital investments, while Dustmann and Glitz analyze how changes in local labor supply skill-mix is absorbed by firms adjusting its production and by exit and entry. Unlike these studies, our analysis focus on firms' R&D investments and on the dynamic effects of immigration.

Gray et al. (2020) also focus on innovation and R&D, and investigate the impact of the EU enlargement on UK firms' spending on process and product innovation. However, unlike this paper, they rely on the standard identification strategy where previous settlement patterns across regions are assumed to predict future settlements. In situations where there is virtually no previous immigration from the relevant countries, this may be a particularly weak instrument. As pointed out above, this was indeed the case in Norway as the Eastern Enlargement happened.

The rest of this paper is organized as follows. Section 2 provides a simple theoretical framework to guide the empirical analysis. Section 3 describes the data, while Section 4 presents the empirical strategy. In Section 5 we report the empirical results and discuss robustness, mechanisms and potential implications for productivity. Section 6 concludes.

2 Theoretical Framework

We set out to investigate if, and how, a large immigration shock has an impact on firms' investments in R&D. To guide the empirical analysis we present a simple theoretical framework similar to that of e.g. Gray et al. (2020). Suppose that there are $i = 1, ..., N$ heterogeneous firms producing output Y_i. Firms are heterogeneous , and organize their production using different occupational mixes. These are assumed to reflect different organizational practices, technology, capital and cultures.

We deviate from Gray et al. (2020) in our focus on occupations rather than regions. Gray et al. (2020) assume there to be a set of distinct regional labor markets, while we assume many distinct occupations, $j = 1, ..., O$, and for each of them a labor market. Labor is thus assumed immobile across occupations. Each firm's production requires the input of several different occupations. Letting L be the supply of labor to one specific occupation, full employment requires that supply is equal to demand and can expressed as

$$L = \sum_{i=1}^{N} C_L^i \left(W, A^i\right) Y^i, \tag{1}$$

where $C_L^i\left(W, A^i\right)$ is the units of occupation o required to produce one unit of output in firm i. Note that keeping productivity and wages constant, labor demand is firm specific reflecting the assumption of firm specific organization of production. A^i is a vector of technology coefficients reflecting firm specific productivity. W is a vector of occupation specific wages. Totally differentiating (1) and re-arranging terms, we obtain

$$dL = \sum_{i=1}^{N} Y^i C_{LW}^i dW + \sum_{i=1}^{N} Y^i C_{LA^i}^i dA^i + \sum_{i=1}^{N} C_L^i dY^i \tag{2}$$

where $C_{LA^i}^i$ measures the changes in unit factor demand induced by changes in the production technology, and C_{LW}^i measures the cross-price effects on factor demands for firm i. Equation (2) states that changes in the supply of labor to a specific occupation must be offset by changes in firms' demand for this occupation generated by some combination of a change in occupation specific wages, technological change, or shift in firms' output. It illustrates that there are in general three margins of adjustment to a positive occupation specific supply shock: (1) through factor prices, (2) through more intensive use of the more abundant occupation within firms, and (3) through an expansion in the size of those firms that use the more abundant occupation more intensively. The first two margins refer to within-firm adjustments, while the third margin refers to between-firm adjustments.

As our empirical analysis of R&D focuses on within-firm adjustments, we shall limit the discussion to this margin. In the short term we can moreover imagine that the technology and organizational practices are fixed, i.e. $dA^i = 0$. Given the time span employed in the empirical analysis, this appears as a reasonable assumption. In this case the within-firm adjustment comes solely from the factor price effect. If immigration induces a shift in labor supply to a given occupation, relative wages will decrease for this occupation as labor becomes more abundant. As a consequence, firms will shift their demand for labor towards these occupations and experience a decline in wage costs. This is the channel along which immigration effects have most frequently been analyzed in the empirical literature and builds on the insight from standard labor market models.

Bratsberg et al. (2019) analyze the impact of the immigration shock to Norway following the Eastern Enlargement of the EU. They find that occupations highly exposed to the immigration shock faced roughly 15 to 25 percent lower wage growth compared to occupations with low exposure over the 2004-2013 period. They also find that the industries most exposed to the immigration shock due to their occupational mix, experienced a relative decline in wage costs.

On this background we set out to investigate how the immigration shock following the Eastern enlargement affected firms' investment in R&D in the short-medium term, and if we can we observe long term consequences with respect to productivity. Our hypothesis is that if immigrants are relatively less skilled in performing R&D, this will lead to a change in relative factor prices across occupations that increases the opportunity cost of engaging in R&D activities and reduces firms' incentives to invest in R&D. We further anticipate that an overall decline in labor costs will reduce firms' incentives to invest in R&D as means of increasing productivity through labor-saving technology.

3 Data

Our empirical analysis of the migration shock is based on four data sets. The first data set is firm level balance sheet data from Statistics Norway for all private non-financial joint-stock companies. The balance sheet data is based on data from annual reports that according to Norwegian law must be filed with the public Register of Company Accounts. The data set contains key account figures related to a firm's income statement and balance sheet.[5]

This panel is matched with Statistics Norway's R&D survey. The survey is based on the OECD Frascati manual[6] and covers the business enterprise sector. It provides biannual information on firm-level R&D investment for a subset of firms. Further details on the R&D

[5]For details on classification and nace codes, see https://www.ssb.no/en/klass/klassifikasjoner/6/versjon/31.
[6]see https://www.oecd.org/sti/inno/frascati-manual.htm

survey are provided in the Appendix Section A. The survey covers all big firms while for smaller firms only a random sample was selected within each strata, defined by NACE 2-digit and firm size class. This feature of the survey, i.e. that the strata is defined by industry and size but not by geography, is common for R&D surveys.

Our third data set is matched employer-employee data, which includes information on wages and occupations as well as immigrant status (country of birth) by person-firm-year. The Norwegian nomenclature (STYRK) for occupations is based on the International Standard Classification of Occupations (ISCO–88) prepared by ILO and further developed by the EU and provides us with 4-digit occupational codes. There are in total 325 occupations in our data.

We have data for the period 2000-2013, while the main period of analysis will be 2003 to 2007. As information about occupations is only available from the year 2003 and onward, this limits the pre-period of our analysis. We merge all three data sources based on a unique firm identifier. The sample covers industries within three main sectors: manufacturing, construction and services. We exclude the primary sector as well as mining and quarrying, which includes the extraction of oil and gas. We also exclude utilities (NACE 35-39) with a majority of publicly owned enterprises, and financial, insurance and real estate activities (NACE 64-68). We drop firms with less than one employee, non-positive wage costs, and/or missing observations for value added. This leaves us with an unbalanced sample of firms where the number of firms varies over the years. Our empirical approach will be based on a difference-in-difference analysis of firms' adjustment at the intensive margin. We therefore require that firms are observed both in the pre- and post-shock period with a positive value of R&D. This limits the sample to a balanced sample of around 800 firms. The fourth data set comes from the O*Net Resource Center and offers detailed information of occupational characteristics. [7] O*Net ranks occupations with respect to a set of requirements. The value 1 means that a given type of skill is not important for the type of work carried out within this occupation, while the value 5 means that it is extremely important. The crosswalk provided by Hoen (2016) allows us to match the O*Net data with the occupational codes used in the Norwegian data. We follow Bratsberg et al. (2019) and use the O*Net data to compute an occupation specific language intensity ($\mathbb{L}_j$). We let it be computed as a simple average of oral and written comprehension and expression requirements, and standardize the variable so that the mean of $\mathbb{L}_j$ is zero and the standard deviation is one.

[7] http://www.onetcenter.org/content.html

4 Empirical Strategy

4.1 Identification

We set out to investigate the effect of the immigration shock to Norway in the aftermath of the EU Eastern enlargement on firms' investment in R&D. As the EU expanded East the Norwegian economy was exposed to an immediate and large migration shock. Over the next four years the number of labor migrants quadrupled and ten years later the employment share of migrants had risen from 7 to 17 percent. Almost 60 percent of the immigrants came from the EU accession countries.

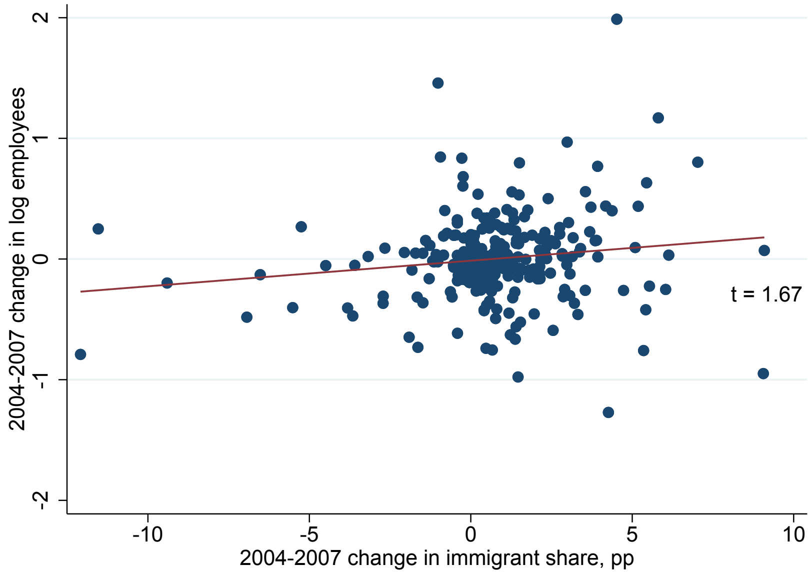
Our empirical strategy is based on two key observations. First, the immigration shock varied substantially across occupations. Second, as firms and industries have different technologies and therefore different occupational mix, they are differently exposed to the immigration shock. Based on these two observations we would expect firms that use occupations with a larger change in immigrant share to be more exposed to the shock.

We follow what is standard in the literature on immigration, see e.g. Borjas (2003), and use change in immigrant share, $\Delta\mu_j$, as our measure of immigrant penetration. Figure 1 illustrates the variation in change in immigrant share by occupation, and the significantly positive relationship between change in overall employment in an occupation and change in immigration share over the period 2004 to 2007.

We compute measures of firm specific exposure to immigration as weighted averages of the change in immigrant shares based on their individual occupational mix. Weights are given by firm specific factor intensities by occupation. Factor intensities are calculated using employer-employee data for the pre-period 2003-2004.[8] To account for differences in human capital across occupations, we use wage shares instead of employment shares when calculating the weights. Our immigration penetration variable is thus defined as $\Lambda_i = \sum_j \lambda_{ij}\Delta\mu_j$, where λ_{ij} depicts the occupation intensity for occupation j in firm i, and $\Delta\mu_j$ depicts change in immigrant share for occupation j.

[8]Factor intensities are based on the average of 2003 and 2004.

Figure 1: Change in Employees versus Change in Immigration Share



Note: The figure shows the percentage point change in immigrantshare on the x-axis versus the percentage point change in total number of employees on the y-axis. The unit of observation is 4-digit STYRK occupation code. The line represents the linear regression line with t-statistics. The sample is constructed based on the same principles as the base sample, but is not limited to employees in firms in the R&D survey. Occupations with less that 25 employees are omitted from the sample.

However, as is well known from analyses on immigration, we face the empirical challenge that the change in the immigrant share may be endogenous. Immigrants may sort into occupations that are intensively used in firms and industries with systematically low or high growth in R&D investments, and this will create a correlation between the error term and the explanatory variable. In order to estimate the causal impact of immigration on R&D investments we therefore need an instrument. To tackle the endogeneity problem many studies have relied on historical settlements across regions to identify effects of immigration. Such approaches exploit the "ethnic enclave" design commonly associated with Altonji and Card (1991). This strategy has nevertheless been subject to criticism. One may argue that the inflow of immigrants is typically autocorrelated, so that the impact of immigration today also captures the longer-term adjustments to previous inflows, see (Jaeger et al., 2018). At the same time the "ethnic enclave" design relies on distinct past settlements of migrants across regions. But due to the East-West divide in Europe, the number of Eastern European migrants to Norway was very low before 2004. In Section B in the Appendix we illustrate the relationship between change immigrant share and past immigrant share in employment across commuting zones (CZ).[9] Irrespective of whether we focus on immigration from the EU8 countries (see Figure 7) or on total immigration (see Figure 8), past settlement turns out to be a poor predictor. In the former case there is no significant relationship, while in the latter case there is actually a significant negative relationship. We therefore choose a different approach that builds on Bratsberg et al. (2019).

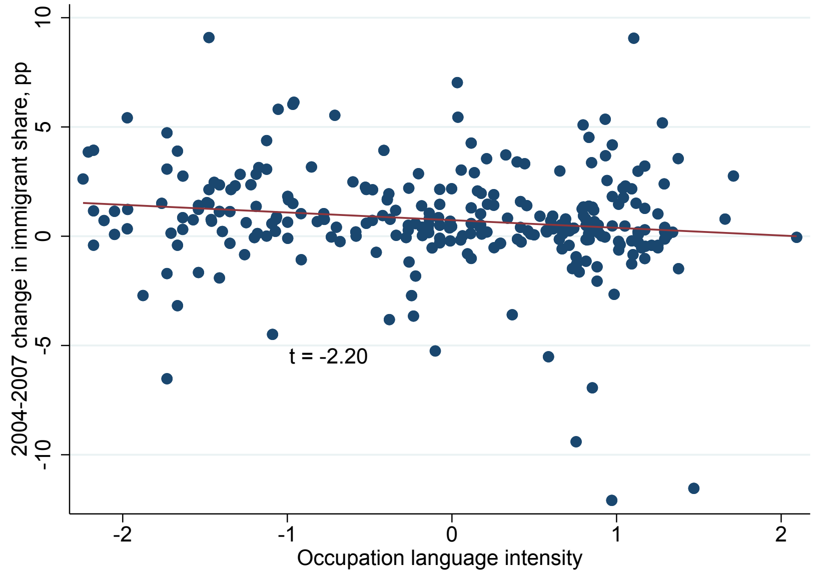
Following Bratsberg et al. (2019) we exploit the fact that occupations differ in terms of language intensity which represents a barrier to foreign workers since Norwegian is the typical workplace language. Based on the O*Net data we compute occupation specific language intensities. We interpret high language requirements as a disutility for immigrants as this requires language training to master the language complexity required for that occupation. Hence, we expect language intensity in occupation j, $\mathbb{L}_j$, to serve as a predictor of the change in immigrant share, $\Delta\mu_j$, and use the former as instrument for the latter.

Since language requirement is time-invariant, our instrument could potentially also be subject to the criticism of the "past settlement" strategy when the inflow of immigrants is autocorrelated, such that the impact of immigration today also captures the longer-term adjustments to previous inflows. However, as the immigration shock we study was instant, comprehensive and externally imposed, we expect this bias to be of minor importance.

We illustrate the relationship between immigrant share and language intensity in Figure 2. In line with what we would expect, the relationship is negative and significant, i.e. occupation with low language intensity experienced higher increase in the immigrant share. Language intensity is thus a powerful predictor of change in immigrant share across occupations.

[9] There are 46 commuting zones in our data set.

Figure 2: Immigrant Share versus Occupation Language Intensity



Note: The figure shows the occupation specific language intensity on the x-axis versus the percentage point change in immigrant share in occupations on the y-axis, where the unit of observation is 4-digit STYRK occupation code. The sample is constructed based on the same principles as the initial sample, but is not limited to employees in firms in the R&D survey. Occupations with less that 25 employees are omitted from the sample.

.

To illustrate the differential exposure to the immigration shock across firms, we use our measure of occupational language intensity to compute the instrument for the firm specific immigration penetration as the negative of the weighted average language intensity of the firm i, i.e. $-\sum_j \lambda_{ij}\mathbb{L}_j$.[10] We then conduct a simple variance decomposition analysis, where we decompose the overall dispersion in firm exposure to immigration into within and between industry components according to $\frac{1}{N_i}\sum_i(\Lambda_i-\overline{\Lambda})^2 = \frac{1}{N_i}\sum_s\sum_{i\in s}(\Lambda_{is}-\overline{\Lambda_s})^2+\frac{1}{N_i}\sum_s N_s(\overline{\Lambda_s}-\overline{\Lambda})^2$, where subscript i refers to firm and subscript s to industry. The results show that about one third of the variation in firms' weighted language intensity is explained by variation between firms within the same industry. This reflects that firms are heterogeneous with respect to their occupational mix even within industries.

4.2 Trends in R&D

Aggregate numbers on R&D in the business sector in Norway show that R&D investments measured as share of value added declined in the period right after the Eastern Enlarge-

[10]We take the negative in order to allow for an easy interpretation of coefficients.

ment.[11] Before we present the empirical model we describe the development in R&D investments based on the firm level data at hand. In line with our identification strategy, we split firms into high and low exposure depending on whether a firm has above or below mean exposure to the immigration shock based on the firm specific measure of language intensity and compute summary measures for the pre-shock and post-shock period. In Tables 1 and 2 we report summary statistics for the unbalanced and balanced sample of firms, see Section 3. Focusing on all firms – irrespective of their exposure – we observe that weighted mean R&D investments (measured as share of value added) declined over time. This is true for the unbalanced as well as the balanced sample, and in line with the aggregate numbers referred to above. The descriptive statistics show that the decline in R&D investments were mainly driven by the firms with a high exposure to immigration.

Table 1: Summary statistics on R&D. Unbalanced sample.

	All firms			High exposure firms			Low exposure firms		
Period	Pre	Post	%Δ	Pre	Post	%Δ	Pre	Post	%Δ
Median	0%	0%	0%	0%	0 %	0%	0%	0%	0%
Weighted Mean	2.52%	2.08%	−18%	1.78%	1.12%	−37%	2.96%	2.81%	−5%
N firms	5363	7138		2834	3963		2529	3175	

Note: R&D investments are measured as R&D expenditure as share of value added.

Table 2: Summary statistics on R&D. Balanced sample

	All firms			High exposure firms			Low exposure firms		
Period	Pre	Post	%Δ	Pre	Post	%Δ	Pre	Post	%Δ
Median	5.89 %	4.02 %	−32%	3.68%	2.48 %	−33%	10.51%	9.20%	−12%
Weighted Mean	4.93%	4.27%	−13%	3.30%	2.25%	−32%	6.33%	6.33%	0%
N firms	837	837		427	427		410	410	

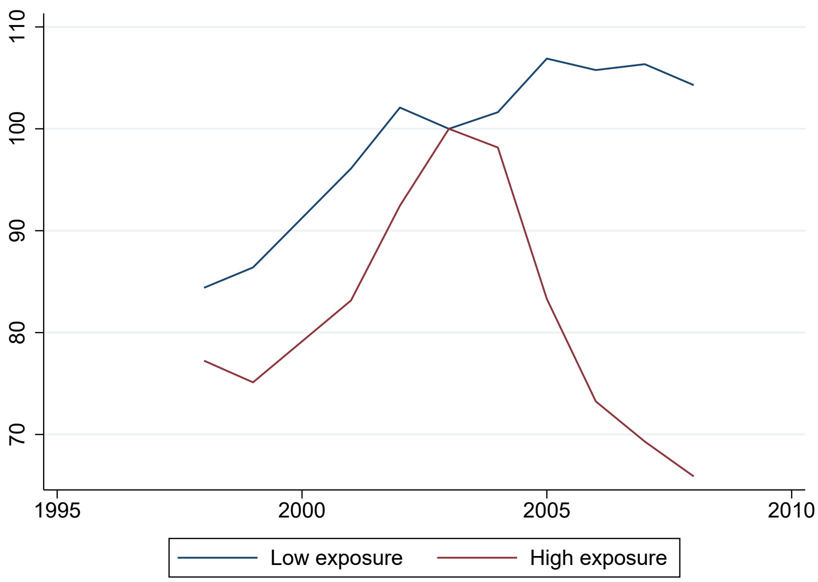
Note: R&D investments are measured as R&D expenditure as share of value added.

Using the balanced panel, which forms the basis for our empirical analysis, we also provide a graphical illustration of the development in median R&D investments using two-year running averages to smooth the curves. Figure 3 illustrates a general increase in median R&D investments in the pre-period for both groups. However, following the EU enlargement

[11] https://stats.oecd.org/

in 2004 there is a decoupling of the time-trend as we compare firms with high and low exposure.[12]

Figure 3: R&D Investments: Low versus High exposure Firms



Note: The unit of observation is weighted mean R&D investments (measured as share of R&D expenditure in value added). Firms are split into two groups – high and low exposure – according to whether their exposure is above or below mean exposure. The figure shows two year running averages normalized to 2003 based on the balanced panel of firms for these two groups.

4.3 The Empirical Model

To analyze the effect of the immigration shock on firms' R&D investments we use a difference-in-difference approach, where we exploit firms' heterogeneous exposure to the shock. We proceed with a standard firm-level difference-in-difference specification given by the linear model:

$$log(RD_{it}) = \alpha_i + \theta_t + \beta(\varLambda_i \times I_t) + \epsilon_i \tag{3}$$

where RD_{it} is firm $i's$ R&D investments in period t. The effect of the immigration shock is captured by the interaction term $(\beta(\varLambda_i \times I_t))$ where Λ_i is the firm exposure to the immigration shock and I_t is a dummy that takes on the value 1 in the shock period and the

[12]Note that graphical illustration in Figure 3 differ from the summary statistics in Table 2, as the former use two years running averages. This explains why the graph suggests a slight increase in R&D for the low exposure firms in the post-period, while the numbers in the Table suggest zero change.

value 0 in the pre-period. θ_t is a year fixed effect and α_i is a firm fixed effect. Differencing (3) across periods yields:

$$\Delta log(RD_i) = log(RD_{i1}) - log(RD_{i0}) = \theta + \beta \mathit{\Lambda}_i + v_i \tag{4}$$

where $t = 0$ denotes the pre-shock period and $t = 1$ denotes the immigration shock period, $\theta \equiv \theta_1 - \theta_0$, and the firm fixed effect α_i is differenced out. The estimated θ will pick up the change in the outcome variable which is due to the business cycle. Firms' exposure ($\mathit{\Lambda}_i$) to the immigration shock is measured by the change in immigration penetration weighted by the firm's occupation mix ($\sum_j \lambda_{ij} \Delta \mu_j$), which in turn is instrumented for by the firm's weighted language intensity ($-\sum_j \lambda_{ij} \mathbb{L}_j$), see Section 4.1.

Variation in Λ_i will allow us to make inference about β. The underlying assumption is that the higherΛ_i, the more exposed a firm is to the immigration shock. Hence, if $\beta < 0$ the immigration shock had a negative impact on the outcome variable, where highly exposed firms (equivalent to low language intensive firms) experienced a relative decrease in the outcome variable. Note, that due to the logs transformation firm observations with zero R&D in both or one of the periods is lost. Hence, our analysis allows us to identify effects at the intensive margin, but not at the extensive margin.

Outcome variable. Our baseline measure of R&D investments will throughout the analysis be a firm's R&D expenditure relative to the firm's value added. However, to ensure robustness, we shall also use other measures of R&D investments.

Sample period. We define the average of the years 2003-2004 as the pre-period and the average of the years 2006-2007 as the shock-period.[13] We choose to use the average over two years in order to avoid measurement errors due to the fact that the R&D survey is biannual. The choice of post-period aims to balance two opposing concerns: we want to allow enough time after the change in migration policy for a migration shock to materialize. At the same time, our analysis is based on the assumption that firms' keep technology and organization, i.e. occupation mix, fixed, which requires that we keep a short to medium term focus. Moreover, the financial crisis hit in 2008, and the both the crisis itself and the different policy measures implemented to combat it, could potentially have affected our results if we had chosen a later post shock period.

Econometric concerns. Our baseline empirical specification includes industry fixed effects. Hence, the identifying variation comes from differences across firms within industries in terms of measured immigration exposure. This specification effectively controls for industry

[13]The EU enlargement happened in May 2004. Ideally we would have liked to choose an earlier pre-period. But data on occupations are not available before 2003. If anything, this will only serve as to bias towards no significant impact.

trends in R&D. A remaining concern is that even within industries differences in immigration exposure are systematically related to other firm characteristics, and that growth in R&D investments is correlated with these characteristics. We address this issue by controlling for pre-shock firm characteristics. Specifically we include pre-shock R&D and pre-shock firm size (number or employees).[14]

However, measured immigration exposure may also be correlated with unobservable firm characteristics which are correlated with R&D growth, due to e.g. technological change and automation. Even in the absence of the immigration shock, we would then expect a significant association between R&D and immigration exposure for other time periods than that of the shock. We address this issue by performing a set of placebo tests choosing pre- and post periods with the same time interval as in the baseline, but for years before and after the shock.

Finally, a potential challenge related to the analysis of R&D investments arises because R&D data are typically based on a survey rather than on an administrative register, where the strata does not account for geography. As a consequence an empirical analysis of R&D investments is not well suited for an approach that relies on exploiting differences across regions in immigration penetration which is common in the literature on immigration, see e.g. Gray et al. (2020). Our empirical strategy which exploits variation across firms in their occupational mix rather than regional variation, obviously overcomes this issue.

5 Empirical Results

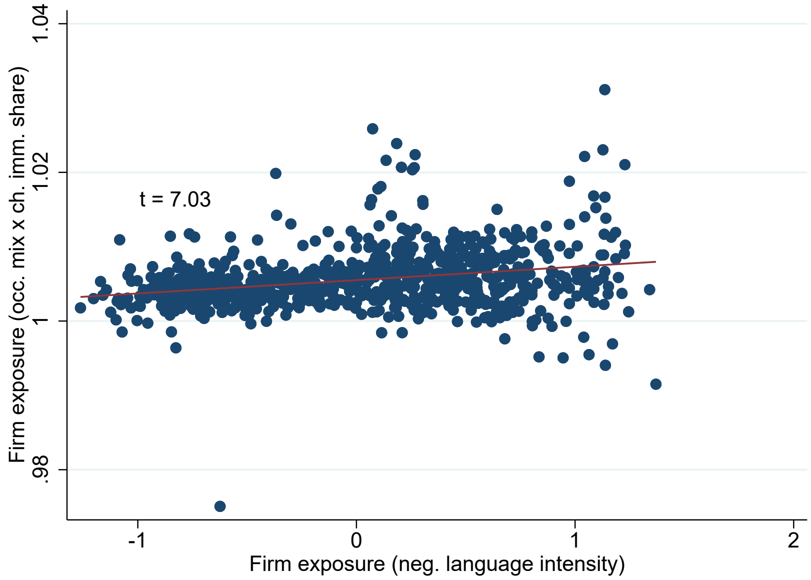
5.1 Immigration and R&D

We proceed by investigating the impact of the immigration shock on firms' R&D investment and estimate the model presented in equation 4 using weighted language intensity as instrument. 2SLS. The sample is the balanced sample of firms.

Figure 4 illustrates the first stage regression, i.e. the relationship between the occupation weighted change in immigrant share and the occupation weighted language intensity. We find that the instrument is strongly correlated with the endogenous variable.

[14] In 2002 the government in Norway introduced a tax incentive in order to encourage investments in R&D, see e.g. Boler et al. (2015). The incentive targeted firms with low investments in R&D. By including pre-shock R&D as a control variable we also account for a potential impact of the tax incentive.

Figure 4: 1st Stage Regression



Note: The figure shows a scatter plot between weighted immigrant penetration ($\sum_j \lambda_{ij} \Delta\mu_j$) on the y-axis and weighted language intensity ($-\sum_j \lambda_{ij} \mathbb{L}_j$) on the horizontal axis. The unit of observation is the firm. Changes refer to the time period 2003/2004 to 2006/2007.

Table 3 shows the 2SLS results as well as first stage results. Column (1) present the estimation results without any controls, while column (2) adds pre-shock firm controls in order to account for differences across firms with respect to technology level (pre-period R&D) and size (pre-period number of employees) and industry-level fixed effects in order to control for aggregate trends. Column (3) adds the same controls but uses 3-digit rather than 2-digit level industry fixed effects. OLS results and reduced form results are provided in Table 6 in the Appendix Section C.

The results suggest that the immigration shock led to a significant decline in the growth of R&D investments in firms which due to their occupational mix were most exposed to the shock. The results are robust to the inclusion of firm controls and industry controls at different levels of aggregation. It appears that the overall decline in R&D investments in the business sector can therefore partly be explained by the immigration shock.

Economic magnitudes. What were the economic magnitudes of the immigration shock on R&D? In order to evaluate the economic impact, we split firms into percentiles according to their exposure to the increase in immigration penetration and compare the 10th and the 90th percentile of firms. We find that the most exposed firms experienced a decline in R&D growth that was more than six times that of the least exposed firms.

Table 3: Immigration and R&D. 2SLS Results.

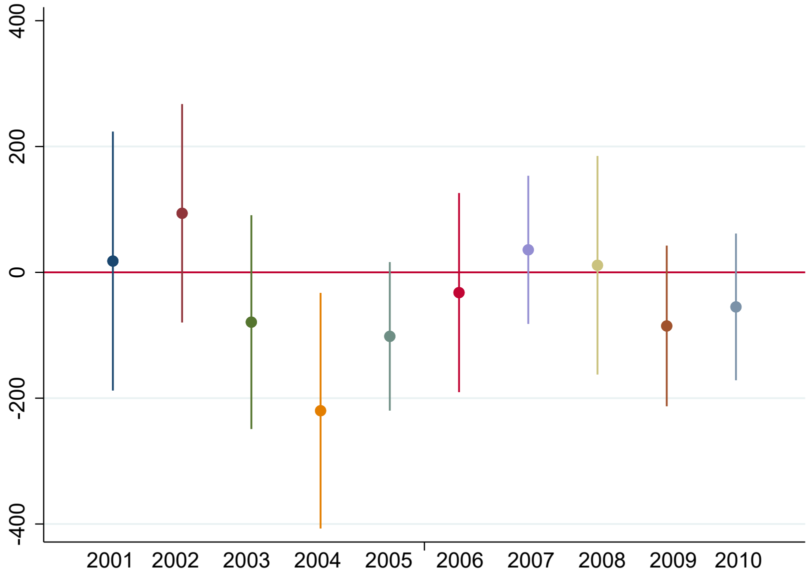
Dependent variable: $\Delta log(RD_i)$	(1)	(2)	(3)
Firm exposure ($\sum_j \lambda_{ij} \Delta\mu_j$)	-93.15^a	-260.3^b	-202.9^b
	(36.58)	(109.0)	(101.2)
Pre-shock Firm controls	No	Yes	Yes
Industry FE (2-digit nace)	No	Yes	No
Industry FE (3-digit nace)	No	No	Yes
	First stage estimates		
$-\sum_j \lambda_{ij} \mathbb{L}_j$	.0018^a	.0015^a	.0015^a
	(.0003)	(.0005)	(.0004)
N	837	837	837

Note: Standard errors clustered by industry in parentheses. The dependent variable is log R&D investments (measured as share of R&D expenditure in value added). Changes refer to the time period 2003/2004 to 2006/2007. The pre-shock firm controls are pre-shock level of the dependent variable and pre-period firm size measured as number of employees. ^{a} p< 0.01, ^{b} p< 0.05, ^{c} p< 0.1.

5.2 Robustness and Mechanisms

Falsification tests. As pointed to in Section 4.3 one potential concern is that there may some underlying trends driving our result, such as technological change and automation. To address this issue, we perform a set of placebo tests choosing pre- and post periods with the same time interval as in the baseline, but for different years. We estimate our preferred baseline model which corresponds to column (3) in Table 7, and illustrate the results in Figure 5. The first point estimate to the left in the figure refers to a regression that uses the average of 2000 and 2001 as pre-period and the average of 2003 and 2004 as post-period, while the point estimate for 2010 refers to a regression that uses the average of 2009 and 2010 as pre-period and the average of 2012 and 2013 as post-period. We see that the only year where there is a significant coefficient is 2004, which refers to the time period we use in our empirical model. For other years before and after the shock, the coefficient is not significant, suggesting that there are no differential general industry-specific trends.

Figure 5: Placebo Tests



Note: The figure provides point estimates for the coefficient of the immigration exposure based on the estimation of equation 4 including firm level pre-shock controls and industry (3-digit nace) fixed effects, for ten different time periods. Vertical lines indicate the 95% confidence intervals. Year given on the horizontal axis refer to the last year of the pre-period.

Alternative dependent variables. To check the robustness of our results we also estimate our baseline specification with two alternative measures of R&D investment. In column (1) in Table 4 we use the level of R&D expenditure as measure, while in column (2) we use the share of R&D employees in total employment. The results confirm the baseline results.

Mechanisms. Our empirical results suggest that the immigration shock to Norway led firms with a high exposure to immigration to reduce its R&D investments relative to firms with a low exposure to immigration. To what extent can these results be ascribed to firm specific exposure, and to what extent is this about exposure that the firm shares with the other firms in the same industry? In order to explore the role of firm versus industry, we estimate a model where we include a measure of immigration exposure computed at the industry rather than the firm level, see column (3) in Table 4. We find that our main explanatory variable, firm specific exposure to immigration, remains significant, while the variable for industry exposure does not appear significant. This means that a firm in a high-exposure industry does not reduce its R&D investment unless this particular firm also has a relatively high exposure. Industry affiliation is in other words a poor predictor of the effect of immigration on a firm's R&D.[15]

[15]In Table 7 in Section D in the Appendix we report empirical results for a regression based on the

Table 4: Immigration and R&D. Robustness. 2SLS Results.

Dependent variable: $\Delta log(RD_i)$	Alternative dep. variable I	Alternative dep. variable II	Industry Exposure
	(1)	(2)	(3)
Immigration exposure ($\sum_j \lambda_{ij}\Delta\mu_j$)	-319.5^b	-219.4^b	-191.6^c
	(131.8)	(100.6)	(99.85)
Immigration exposure - industry level (3-digit nace)			-47.26
			(86.70)
Pre-shock Firm controls	Yes	Yes	Yes
Industry FE (3-digit nace)	Yes	Yes	No
N	838	838	837

Note: Standard errors clustered by industry in parentheses. The dependent variable in the column (1) is the change in log of the level of R&D expenditure. Changes refer to the time period 2003/2004 to 2006/2007. The dependent variable in column (3) is the change in log of full-time equivalent R&D employees as share of total employees. The pre-shock firm controls are pre-shock level of the dependent variable and pre-period firm size measured as number of employees. ^{a} p< 0.01, ^{b} p< 0.05, ^{c} p< 0.1.

5.3 Implications for Productivity

Theoretical and empirical research has documented a strong positive impact of the role of investments in R&D on firms' productivity. Hence, we would expect that any negative shock to R&D would consequently also have a negative impact on firms' productivity. Having established a negative impact of the immigration shock on R&D we proceed by investigating what happened to firm productivity in the short and long run. We estimate revenue based total factor productivity (TFP) using the Levinsohn-Petrin approach (Levinsohn and Petrin, 2003). For more details on the estimation, see the Appendix Section E.

We proceed by backing out firm level TFP and apply the same empirical approach as above to examine the short and long run effects of the immigration shock on productivity:

$$\Delta log(TFP_i) = log(TFP_{1t}) - log(TFP_{0t}) = \theta + \beta\Lambda_i + \boldsymbol{\phi}\mathbf{x}_{i0} + v_i \quad (5)$$

where $t = 0$ denotes the pre-shock period and $t = 1$ denotes the immigration shock period, $\theta \equiv \theta_1 - \theta_0$, and the firm fixed effect α_i is differenced out. The estimated θ picks

specification in (4) where we use industry level data. We see that more exposed industries experienced a significantly higher reduction in R&D growth. Combined with our firm level results, we observe that these results are driven by a difference in exposure across firms within industries.

up the change in the outcome variable which is due to the business cycle. We estimate (5) using 2SLS, and report the results in Table 5 using four different post-periods, while keeping 2003/2004 as pre-period throughout. The first-stage is identical to the column (4) in Table 3. Column (1) shows the short run effect on TFP using 2006/2007 as the post-period. In line with what we would expect given the presumed lagged effect of R&D on productivity growth, the results suggest that there was no productivity effects of immigration in the short run. However, in the medium to long run after the initial shock, immigration appears to have had a significant negative impact on TFP, see columns (4)-(6) in Table 5.

Table 5: Effect of immigration shock on TFP. 2SLS results.

Dependent variable: $\Delta log(TFP_i)$	(1)	(2)	(3)	(4)	(5)	(6)
Post periods	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012
Immigration exposure $(-\sum_j \lambda_{ij}\mathbb{L}_j)$	-37.84	-28.06	-76.76	-120.2^c	-239.7^b	-257.7^a
	(43.03)	(48.29)	(47.39)	(65.13)	(104.8)	(93.01)
Pre-shock firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE (3-digit nace)	Yes	Yes	Yes	Yes	Yes	Yes
N	816	709	610	571	610	571

Note: The dependent variable is $\Delta log(TFP_{si})$. Firm controls are pre-shock log TFP and log number of employees. Industry controls are pre-shock log skill intensity and log capital intensity. The pre-shock firm controls are pre-shock level of the dependent variable and pre-period firm size measured as number of employees. Standard errors in parentheses. ^{a} p< 0.01, ^{b} p< 0.05, ^{c} p< 0.1.

Figure 6 illustrates the development in the impact of the immigration shock over time. We show the estimated coefficients from table 5 with 95% confidence intervals. The figure illustrates that there is no immediate effect in the aftermath of the shock for high relative to low exposure firms, but an increasingly negative effect on productivity for high relative to low exposure firms over time. Our analysis does not allow us to conclude on a causal relationship between the immigration induced decline in R&D investments and the reduced productivity growth, but the combined results on R&D investments and TFP do indicate such a link.

Figure 6: Effect of immigration on TFP year-by-year

Note: The figure provides point estimates for the coefficient of the immigration exposure based on the estimation of equation 5 including firm level pre-shock controls and industry (3-digit nace) fixed effects, for the six different time periods reported in table 5. Vertical lines indicate the 95% confidence intervals. Year given on the horizontal axis refer to the post-period.

6 Conclusions

There is an extensive literature on the effect of immigration on the economy through its effect on workers, wages and employment. The empirical evidence on the impact immigration on firm performance and particularly on R&D is, however, scant. In this paper we investigate the impact of the major immigration shock to Norway after the EU enlargement in 2004 on firms' investment in R&D. We find that the immigration shock led to a reduction in firms' R&D. An overall decline in R&D investments in the business sector over these years can therefore partly be explained by the increase in immigration penetration.

We also investigate the long term consequences of the immigration shock for productivity. In line with what we would expect given the well documented relationship between R&D and productivity, our results indicate that the shock also led to reduced productivity in the long run.

We note that a previous study, see Gray et al. (2020), finds a positive relationship between the Eastern enlargement immigration shock and firms' innovation activity. Our results are robust to a set of econometric issues and stand in distinct contrast to their results. The empirical evidence presented in this paper suggests that the flip side of a vast increase in labor supply, may be a reduced incentive to conduct R&D with long term negative consequences for productivity.

References

Acemoglu, Daron, "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics*, 1998, *113*, 1055–1089.

Altonji, Joseph and David Card, "The Effect of Immigration on the Labor Market Outcomes of Less-Skilled Natives," in John M. Abowd and Richard B. Freeman, eds., *Immigration, Trade and the Labor Market*, University of Chicago Press, Chicago, 1991.

Boler, Esther Ann, Andreas Moxnes, and Karen Helene Ulltveit-Moe, "R&D, International Sourcing and the Joint Impact on Firm Performance," *American Economic Review*, 2015, *105* (12), 3704–3739.

Borjas, George J., "The labor demand curve is downward sloping: Reexamining the impact of immigration on the local labor market," *The Ouarterly Journal of Economics*, 2003, *118* (4), 1335–1374.

Bratsberg, Bernt, Andreas Moxnes, Oddbjorn Raaum, and Karen Helene Ulltveit-Moe, "Opening the Floodgates: Industry and Occupation Adjustments to Labour Migration," *CEPR Discussion Paper No. 13670*, 2019.

Dølvik, Jon Erik and Line Eldring, *Mobility of labour from new EU states to the Nordic Region: – Development trends and consequences*, Nordic Council of Ministers, 2008.

Dustmann, Christian and Albrecht Glitz, "How do Industries and Firms Respond to Changes in Local Labor Supply?," *Journal of Labor Economics*, 2015, *33* (3), 711–750.

Gray, Rowena, Guilia Montresor, and Greg C. Wright, "Processing Immigration Shocks: Firm Repsonses on the Innovation Margin," *Journal of International Economics*, 2020, *126.*

Hoen, Maria Brunborg, "Occupational crosswalk, data and language requirements," *Frisch Centre Working Paper 1/2016*, 2016.

Jaeger, David A., Joakim Ruist, and Jan Stuhler, "Shift-Share Instruments and the Impact of Immigration," 2018.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70* (2), 317–342.

Lewis, Ethan G., "Immigration, skill mix and capital skill complementarity," *The Quarterly Journal of Economics*, 2011, *126*, 1029–1069.

Peri, Giovanni and Chad Sparber, "Task Specialization, Immigration and Wages," *American Economic Journal: Applied Economics*, 2009, *1* (3), 135–169.

Shu, Pian and Claudia Steinwender, "The impact of trade liberalization on firm productivity and innovation," in "NBER Book Series Innovation Policy and the Economy," Vol. 19, Josh Lerner and Scott Stern, 2019.

Stuen, Eric, Ahmed Mushfiq Mobarak, and Keith Maskus, "Skilled Immigration and Innovation: Evidence from Enrolment Fluctuations in US Doctoral Programmes," *Economic Journal*, 2012, *122*, 1143–1176.

Appendix

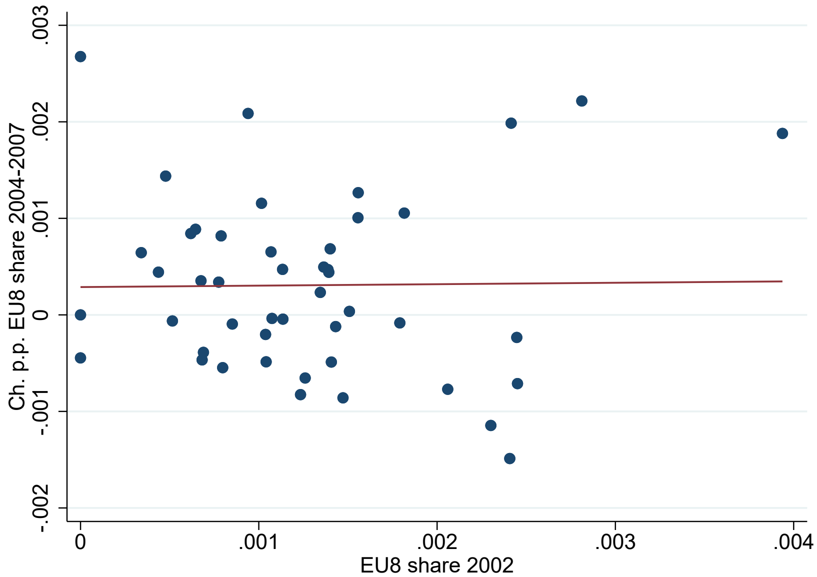
A R&D Data

The R&D survey measures R&D activity in the Norwegian business enterprise sector.[16] The survey is conducted based on the the Frascati manual (see https://www.oecd.org/sti/inno/frascati-manual.htm) and the statistics are comparable to statistics for other countries and are reported to the OECD and EUROSTAT. The R&D survey includes: (i) all firms with at least 50 employees; (ii) all firms with less than 50 employees and with reported intramural R&D activity in the previous survey of more than NOK 1 million or extramural R&D of more than NOK 3 million; (iii) among other firms with 10-49 employees a random sample was selected within each strata (NACE 2-digit and firm size class). The survey covers the business enterprise sectors and includes the following NACE industries: 03, 05-33, 35-39, 41-43, 46, 49-53, 58-66, 70-72, 74.9 and 82.9.

[16] It includes the entire manufacturing sector and the majority of the service sector, but leaves out some service industries with insignificant R&D activity.

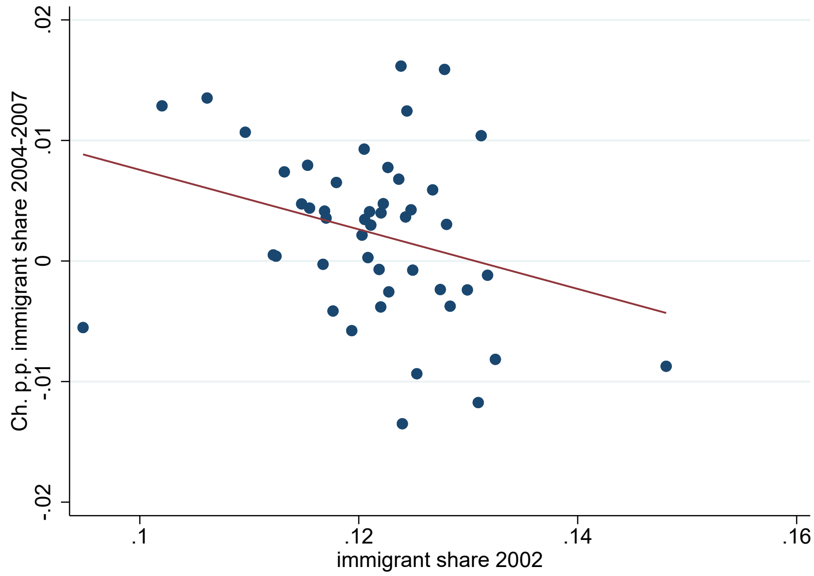
B Alternative Instruments based on Past Settlements

Figure 7: Change in EU8 Immigrant Share versus EU8 past Immigrant Share by Region (CZ)



Note: The figure shows the percentage point change in EU8 immigrant share (2004-2007) versus past EU8 immigrant share (2002) in employment. Unit of observation is commuting zone (CZ). EU8 refers to the eight Eastern European countries that became members of the EU in 2004. The sample is constructed based on the same principles as the initial sample, but is not limited to employees in firms in the R&D survey.

Figure 8: Change in Immigrant Share versus past Immigrant Share by Region (CZ)



Note: The figure shows the percentage point change in immigrant share (2004-2007) versus past immigrant share (2002) in employment. Unit of observation is commuting zone (CZ). by The sample is constructed based on the same principles as the initial sample, but is not limited to employees in firms in the R&D survey.

C OLS & Reduced Form Results

Table 6: Immigration shock and R&D. OLS & Reduced Form Results.

Dependent variable $\Delta log(RD_i)$	OLS	Reduced Form
Immigration exposure ($\sum_j \lambda_{ij}\Delta\mu_j$)	10.57	-.313^a
	(15.90)	(.107)
Pre-shock Firm controls	Yes	Yes
Industry FE (3-digit nace)	Yes	Yes
N	837	837

Note: Standard errors clustered by industry in parentheses. The dependent variable is log R&D investments (measured as share of R&D expenditure in value added). Changes refer to the time period 2003/2004 to 2006/2007. The pre-shock firm controls are pre-shock level of the dependent variable and pre-period firm size measured as number of employees. Robust standard errors in parentheses. ^{a} p< 0.01, ^{b} p< 0.05, ^{c} p< 0.1.

D Industry Level Results

Table 7: Immigration shock and R&D intensity. Industry level. 2SLS Results.

Dependent variable	$\Delta log(RD_s)$
Industry exposure ($\sum_j \lambda_{sj} \Delta\mu_j$)	-147.7^b
	(71.6)
Industry Pre-shock controls	Yes
	First stage
$-\sum_j \lambda_{sj} \mathbb{L}_j$	0.0029^a
	(.0008)
N	100

Note: The dependent variable is log R&D intensity (measured as R&D expenditure in value added) at the 3-digit industry level. Changes refer to the time period 2003/2004 to 2006/2007. Industry controls are pre-shock log skill intensity and log capital intensity. Robust standard errors in parentheses. ^{a} p< 0.01, ^{b} p< 0.05, ^{c} p< 0.1.

E Productivity Estimation

Our point of departure is a production function that takes the form

$$y_{ist} = \beta_0 + \beta_{l,s} l_{ist} + \beta_{k,s} k_{ist} + \omega_{ist} + \epsilon_{ist}$$

where y_{ist} denotes log value added of firm i in industry s at time t, l_{ist} is log number of employees and $k_{ist}$is the log capital stock. The output elasticities with respect to labor and capital are $\beta_{l,s}$ and $\beta_{k,s}$, respectively. The unobserved, hicks-neutral productivity is denoted by ω_{ist}, and ϵ_{ist} depicts the error term. We estimate the production function using a structural proxy estimator. We follow Levinsohn and Petrin (2003) and use intermediate inputs as proxy for the productivity term ω_{ist}. Note that our results are robust to different measures of labor and to the use of Olley-Pakes rather than Levinsohn-Petrin.[17]to the particular choice for production function method. We use the estimated elasticities $\widehat{\beta_{l,s}}$ and $\widehat{\beta_{k,s}}$ and back out the log TFP (tfp_{ist})for firm i in industry s at time t as:

$$tfp_{ist} = \omega_{ist} = y_{ist} - \widehat{\beta_{l,s}} \cdot l_{ist} - \widehat{\beta_{k,s}} \cdot k_{ist}$$

[17]We have estimated the production function using wagebill rather than number of employees and split employees into skilled and unskilled.